

CURRENT APPOINTMENT

Bioinformatics Scientist

Faculty of Arts and Sciences
Harvard University, Cambridge, MA

2021 – present

EDUCATION

Distinguished Doctor of Philosophy Indiana University Bloomington, IN	Bioinformatics and Evolutionary Biology	2019
Master of Science Indiana University Bloomington, IN	Bioinformatics	2013
Bachelor of Science Purdue University West Lafayette, IN	Biology	2010

PROFESSIONAL APPOINTMENTS

Postdoctoral Research Associate Division of Biological Sciences University of Montana, Missoula, MT	2019 – 2021
---	-------------

PUBLICATIONS

** = undergraduate mentee

1. Yan H, Hu Z, **Thomas GWC**, Edwards SV, Sackton TB, Liu JS. 2023. PhyloAcc-GT: A Bayesian method for inferring patterns of substitution rate shifts on targeted lineages accounting for gene tree discordance. *Molecular Biology and Evolution*. 40(9):msad195
2. Moore EC, **Thomas GWC**, Mortimer S, Kopania EEK, Hunnicutt KE, Clare-Salzler ZJ, Larson EL, Good JM. (2022). The evolution of widespread recombination suppression on the dwarf hamster (*Phodopus sungorus*) X chromosome. *Genome Biology and Evolution*. 14(6):evac080.
3. **Thomas GWC**, Wang RJ, Nguyen J**, Harris RA, Raveendran M, Rogers J, Hahn MW. (2021). Origins and long-term patterns of copy-number variation in rhesus macaques. *Molecular Biology and Evolution*. 38(4):1460-1471.

4. Sun C, ..., **Thomas GWC**, ..., Mueller RL. 2021. Genus-wide characterization of bumblebee genomes reveals variation associated with key ecological and behavioral traits of pollinators. *Molecular Biology and Evolution*. 38(2):486-501.
5. Wang RJ, **Thomas GWC**, Raveendran M, Harris RA, Doddapaneni H, Muzny DM, Capitanio JP, Radivojac P, Rogers J, Hahn MW. 2020. Paternal age in rhesus macaques is positively associated with germline mutation accumulation but not with measures of offspring sociability. *Genome Research*. 30:826-834.
6. **Thomas GWC**, Dohmen E, Hughes ST, Murali SC, Poelechau M, Glastad K, ..., Chipman AD, Waterhouse RM, Bornberg-Bauer E, Hahn MW, Richards S. 2020. The genomic basis of Arthropod diversity. *Genome Biology*. 21(15).
7. Bentz AB, **Thomas GWC**, Rusch DB, Rosvall KA. 2019. Tissue-specific expression profiles and positive selection analysis in the tree swallow (*Tachycineta bicolor*) using a *de novo* transcriptome assembly. *Scientific Reports*. 9:15849.
8. **Thomas GWC** and Hahn MW. 2019. Referee: reference assembly quality scores. *Genome Biology and Evolution*. 11(5):1483-1486.
9. Rogers J, ..., **Thomas GWC**, ..., Jolly CJ, Gibbs RA, Worley KC. 2019. The comparative genomics and complex population history of *Papio* baboons. *Science Advances*. 5(1).
10. Da Lage J-L, **Thomas GWC**, Bonneau M, Courtier-Orgogozo V. 2019. Evolution of salivary glue genes in *Drosophila* species. *BMC Evolutionary Biology*. 19(36).
11. Prost S, Armstrong EE, Nylander J, **Thomas GWC**, Suh A, Petersen B, Dalen L, Benz BW, Blom MPK, Palkopoulou E, Ericson PGP, Irestedt M. 2019. Comparative analyses identify genomic features potentially involved in the evolution of birds-of-paradise. *GigaScience*. 8(5).
12. **Thomas GWC**, Wang RJ, Puri A**, Harris RA, Raveendran, Hughes DST, Murali SC, Williams LE, Doddapaneni, Muzny DM, Gibbs RA, Abee CR, Galinski MR, Worley KC, Rogers J, Radivojac P, Hahn MW. 2018. Reproductive longevity predicts mutation rates in primates. *Current Biology*. 28(19):3193-3197.
13. Warren WC, García-Pérez R, ..., **Thomas GWC**, ..., Schartl M. 2018. Clonal polymorphism and high heterozygosity in the celibate genome of the Amazon molly. *Nature Ecology and Evolution*. 2:669-679.
14. Schoville SD, Chen YH, ..., **Thomas GWC**, ..., Richards S. 2018. A model species for agricultural pest genomics: the genome of the Colorado potato beetle, *Leptinotarsa decemlineata* (Coleoptera: Chrysomelidae). *Scientific Reports*. 8(1931).
15. Palesch D, Bosinger SE, ..., **Thomas GWC**, ..., Silvestri G. 2018. Sooty mangabey genome sequence provides insight into AIDS resistance in a natural SIV host. *Nature*. 553:77-81.

16. **Thomas GWC**, Ather SA**, and Hahn MW. 2017. Gene-tree reconciliation with MUL-trees to resolve polyploidy events. *Systematic Biology*. 66(6):1007-1018.
17. **Thomas GWC**, Hahn MW, and Hahn Y. 2017. The effects of increasing the number of taxa on inferences of molecular convergence. *Genome Biology and Evolution*. 9(1):213-221.
18. Warren WC, ..., **Thomas GWC**, ..., Freimer NB. 2015. The genome of the vervet (*Chlorocebus aethiops sabaesus*). *Genome Research*. 25(12):1921-1933.
19. **Thomas GWC** and Hahn MW. 2015. Determining the null model for detecting adaptive convergence from genomic data: a case study using echolocating mammals. *Molecular Biology and Evolution*. 32(5):1232-1236.
20. Foote AD, Liu Y, **Thomas GWC**, Vinař T, ..., Gibbs RA. 2015. Convergent evolution of the genomes of marine mammals. *Nature Genetics*. 47(3):272-275.
21. Neafsey DE, Waterhouse RM, ..., **Thomas GWC**, ..., Besansky NJ. 2014. Highly evolvable malaria vectors: The genomes of 16 Anopheles mosquitoes. *Science*. 347.
22. Montague MJ, ..., **Thomas GWC**, ... Warren WC. 2014. Comparative analysis of the domestic cat genome reveals genetic signatures underlying feline biology and domestication. *Proc Natl Acad Sci USA*. 111(48):17230-17235.
23. Carbone L, ... **Thomas GWC**, ..., Gibbs RA. 2014. Gibbon genome and the fast karyotype evolution of small apes. *Nature*. 513:195-201.
24. **Thomas GWC** and Hahn MW. 2014. The human mutation rate is increasing, even as it slows. *Molecular Biology and Evolution*. 31(2):253-257.
25. Han MV, **Thomas GWC**, Lugo-Martinez J, and Hahn MW. 2013. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Molecular Biology and Evolution*. 30(8):1987-1997.

PRE-PRINTS

1. Kopania EEK, **Thomas GWC**, Hutter CR, Mortimer SME, Callahan CM, Roycroft E, Achmadi AS, Breed WG, Clark NL, Esselstyn JA, Rowe KC, Good JM. Molecular evolution of male reproduction across species with highly divergent sperm morphology in diverse murine rodents. *bioRxiv*.
2. **Thomas GWC**, Hughes JJ, Kumon T, Berv JS, Nordgren CE, Lampson M, Levine M, Searle JB, Good JM. The genomic landscape, causes, and consequences of extensive phylogenomic discordance in Old World mice and rats. *bioRxiv*.
3. Mirchandani C, Shultz AJ, **Thomas GWC**, Smith SJ, Baylis M, Arnold B, Corbett-Detig R, Enbody E, Sackton TB. A fast, reproducible, high-throughput variant calling workflow for evolutionary, ecological, and conservation genomics. *bioRxiv*.

PRESENTATIONS & POSTERS

- 1. Quantifying and mitigating reference bias in comparative genomics**
Evolution Meeting, Albuquerque, NM
Contributed talk
2023

- 2. Prioritizing loci for ILS-aware rate analyses using phylogenetic concordance factors**
Society of Molecular Biology and Evolution Global Symposium 2 (Sustainability, Equity, and Efficiency in Computational Biology), virtual
Contributed talk
2022
- 3. Molecular and morphological evolution across the most species-rich radiation in mammals**
Evolution Meeting, Cleveland, OH
Poster
2022
- 4. Prioritizing loci for ILS-aware rate analyses using phylogenetic concordance factors**
Evolution Meeting, Cleveland, OH
Contributed talk
2022
- 5. Speciation and introgression across the most species-rich radiation in mammals**
Population, Evolutionary, and Quantitative Genetics Conference, Pacific Grove, CA
Platform talk
2022
- 6. Patterns of genomic variation across the tree of life**
University of Massachusetts Lowell Department of Biology Seminar, Lowell, MA
Invited talk
2022
- 7. Patterns of genomic variation across the tree of life**
Harvard Museum of Comparative Zoology Lunchtime Seminar, Cambridge, MA
Invited talk
2022
- 8. Pedigree sequencing and mutation rate variation in primates**
Meeting of the American Association of Biological Anthropologists, virtual
Invited talk
2021
- 9. Origins and long-term patterns of genomic variation across the tree of life**
Harvard University, Bioinformatics group, virtual
Invited talk
2021
- 10. Origins and long-term patterns of genomic variation across the tree of life**
Binghamton University, Department of Biological Sciences, virtual
Invited talk
2021

11. **The origins and patterns of genomic variation across the tree of life**
Utah Valley University, Department of Biology, virtual
Invited talk 2021

12. **Patterns of molecular evolution in Arthropods**
Arthropod Genomics Symposium, virtual
Invited talk 2020

13. **Causes and consequences of structural variation in the *Macaca mulatta* genome**
First author Nguyen J**
Center of Excellence for Women & Technology Research Experience for Undergraduates Symposium, Bloomington, IN
Poster 2019

14. **Reproductive longevity predicts mutation rates in primates**
Population, Evolutionary, and Quantitative Genetics Conference, Madison, WI
Platform talk 2018

15. **The evolution of the genes and genomes of 76 arthropod species**
Evolution Meeting, Portland, OR
Regular talk 2017

16. **The evolution of the genes and genomes of 76 arthropod species**
Arthropod Genomics Symposium, Notre Dame University, South Bend, IN
Invited talk 2017

17. **Gene-tree reconciliation with MUL-trees for polyploidy analysis**
Evolution Meeting, Austin, TX
Regular talk 2016

18. **Accounting for sequencing error in phylogenetics**
Society of Systematic Biologists, University of Michigan, Ann Arbor, MI
Lightning Talk 2015

19. **Inferring molecular convergence from genomic data**
Midwest Ecology and Evolution Conference, Indiana University, Bloomington, IN
Contributed talk 2015

20. **Convergent evolution of the genomes of marine mammals**
Society for Molecular Biology and Evolution, San Juan, Puerto Rico
Contributed talk 2014

21. **Convergent evolution of the genomes of marine mammals**
Genetics, Cellular, and Molecular Sciences Symposium, Bloomington, IN
Poster 2014

RESEARCH EXPERIENCE

Bioinformatics Scientist

FAS Informatics & Scientific Applications Group
Harvard University, Cambridge, MA

2021 – present

- Joined a team developing methods to account for phylogenetic discordance during Bayesian estimation of substitution rates (PhyloAcc).
- Develop genomic and phylogenetic software with a main goal for reproducible and accurate inference, including Snakemake pipelines for whole genome alignment using GPU accelerated compute nodes and annotation of degeneracy of coding transcripts (degenotate).
- Design and teach workshops on bioinformatics and programming for graduate students at the university.
- Assist researchers from across the university with genomic and phylogenetic analyses.
- Teach workshops on how to use Bayesian substitution rate estimation software (PhyloAcc).
- Present work to researchers outside the university at conferences regarding evolutionary genomics and bioinformatics.

Postdoctoral Research Associate

Laboratory of Jeffrey Good
Division of Biological Sciences
University of Montana, Missoula, MT

2019 – 2021

- Lead a comparative project to study molecular evolution and phylogenetics in murine rodents using whole exome sequences from 210 species and whole genome sequences from dozens of species.
- Devised best-practices for assembly and annotation of a large sample of exomes.
- Update and maintain software released by the lab to automate reference-guided genome assembly through iterative mapping (pseudo-it).
- Analyze the phylogenetic relationships of newly sequenced rodent genomes using the reference genomes and genetic resources from the well annotated mouse and rat systems to build an empirical landscape of phylogenetic discordance across chromosomes.
- Applied for grants (NIH NRSA) to expand whole genome sampling of murine rodents to study phylogenetic discordance, patterns of molecular evolution, convergent evolution, and sex chromosome rearrangements.
- Administrator of lab's computational resources, including two 32 core 200GB servers and one 192TB NAS, and the lab github account.
- Mentor and guide graduate and undergraduate students in the lab regarding computational methods in genomics, phylogenetics, and molecular evolution.

Research Assistant

Laboratory of Matthew Hahn
School of Informatics, Computing, and Engineering
Department of Biology
Indiana University, Bloomington, IN

2012 – 2019

- Developed a method to estimate genome assembly and annotation error from gene count data using CAFE's error model function (caferror).

- Studied patterns of convergent evolution in marine mammals and echolocating mammals and devised best practices for identifying molecular convergence.
- Devised a method to infer the presence and mode of polyploidy from gene tree topologies (GRAMPA).
- Modeled and observed mutation rate patterns in primates, including single nucleotide mutations and structural variants, by sequencing families of owl monkeys and macaques.
- Led the comparative phylogenetic portion of the i5K pilot project which involved analyzing the genomes of 76 arthropods.
- Wrote software to annotate genomes with quality scores (Referee).
- Participated in several collaborations by performing comparative analyses, such as phylogeny reconstruction and assessment, gene family analysis, and positive selection scans.

TEACHING EXPERIENCE

Instructor

Informatics Group workshops
Harvard University

2023

- Developed and taught workshops to biology grad students and postdoctoral fellows.
- Designed hands-on activities in R markdown to demonstrate basic R syntax, data manipulation with the tidyverse, and plotting with ggplot.
- Designed hands-on activities in R markdown to teach about command line tools like awk, grep, samtools, bedtools, and bcftools.
- Introduced basic bioinformatics file formats to students, such as fasta, fastq, bed, bam, gff, and vcf files.
- Demonstrated best practices in bioinformatics.

Guest Instructor

OEB 275R: Comparative Genomics: Phylogenetic Approaches to Linking
Genomes and Phenotypes
Prof. Scott Edwards
Harvard University

2022

- Led graduate students in discussions on comparative genomics and bioinformatics
- Designed hands-on activities to demonstrate how to infer accelerated substitution rates on a phylogeny using tools such as PhyloAcc, git, and RStudio and R markdown.
- Guided the students through a web-based workshop on how to use Bayesian substitution rate estimation software (PhyloAcc).
- Demonstrated best practices in bioinformatics.

Instructor

Conservation Genetics and Population Genomics course (ConGen)
Virtual course
University of Montana

2020 – 2023

- Gave keynote lecture on genome sequencing and assembly.
- Designed hands-on activities for a 2-hour workshop on genome assembly and read mapping.
- Designed and presented a workshop to teach introductory bioinformatics skills including project organization, common bioinformatics file formats, and examples of basic bioinformatics tasks

- Met one-on-one with students during office hours to discuss and give advice about their data.

Student Mentor

School of Informatics, Computing, and Engineering, 2014 – 2019
 Department of Biology
 Indiana University, Bloomington, IN

Provided guidance to high school and undergraduate students in conceptualizing evolution by involving them in various computational projects, providing a basis in programming, data analysis, and scholarship.

- CEWiT Research Experience for Undergraduate Women 2018 – 2019
- Computer Science Independent Study 2017 – 2018
- Computer Science Independent Study 2016 – 2017
- Jim Holland Summer Science Research Program 2014

Teaching Assistant

School of Informatics, Computing, and Engineering, 2011 – 2016
 Department of Biology
 Indiana University, Bloomington, IN

Taught lab sessions, led class discussions, graded assignments, and met with students individually to assist them.

- INFO-I211: Information Infrastructure 2014, 2016
- BIOL-Z620/INFO-I590: SNP Discovery and Population Genetics 2014
- INFO-I308: Information Representation 2011 – 2012

PROFESSIONAL SERVICE

Graduate Student Advisor

Indiana University Bioinformatics Club 2012 – 2014
 Indiana University, Bloomington, IN

Served as a co-founding member and treasurer (2012 only) to raise awareness of bioinformatics and associated opportunities for undergraduate and graduate students by facilitating group projects and discussions, tours, and social events.

Peer Review

- *G3* (3x)
- *PLoS ONE*
- *Molecular Biology and Evolution* (2x)
- *New Phytologist*
- *Pacific Symposium of Biocomputing* (2x)
- *Genes* (2x)
- *Nature Communications* (2x)
- *Systematic Biology* (3x)
- *Evolution Letters*
- *Science Advances*
- *Genome Biology and Evolution* (5x)
- *Society of Systematic Biologists Graduate Student Research Award* (2x)
- *Molecular Ecology Resources* (2x)
- *Molecular Ecology* (2x)
- *BMC Ecology and Evolution*

- *Genomics*
- *GigaScience*

- *PeerJ*
- *BMC Biology*

GRANTS

Genetics, Cellular, and Molecular Sciences Training Grant

Department of Biology
Indiana University, Bloomington, IN

2014 – 2015

Sandy Ostroy Summer Research Award for Undergraduates

Department of Biology
Purdue University, West Lafayette, IN

2008

AWARDS

Distinguished Ph.D. Dissertation Award

The University Graduate School
Indiana University, Bloomington, IN

2020

SOFTWARE & RESOURCES

Bioinformatics tips & tricks workshop

<https://harvardinformatics.github.io/workshops/2023-spring/biotips/>

- A workshop our group developed to teach the basics of command line tools and file formats used in bioinformatics

Introduction to R workshop

<https://harvardinformatics.github.io/workshops/2023-spring/r/>

- A workshop our group developed to teach introductory data analysis in R, including the tidyverse and plotting with ggplot

PhyloAcc workshop

<https://gwct.github.io/phyloacc/OEB275R/index.html>

- A website I designed to guide students through the PhyloAcc workshop during my guest lecture in OEB275R (Comparative genomics).

PhyloAcc: Bayesian estimation of substitution rates while accounting for phylogenetic discordance

<https://phyloacc.github.io/>

- Joined development team to develop methods to speed up inferences and improve usability of software.

degenotate: Annotation of coding sites with codon degeneracy and MK tables

<https://github.com/harvardinformatics/degenotate>

- This software outputs a bed file containing information about the degeneracy of every coding site in a genome as well as counts of polymorphisms for MK tests.

bonsai: Tree pruning with concordance factors

<https://github.com/gwct/bonsai>

- Prunes large phylogenetic trees to maximize the concordance of the underlying alignments.

pseudo-it: Pseudo-genome assembly with iterative mapping

<https://github.com/goodest-goodlab/pseudo-it>

- This software iteratively maps reads to generate a pseudo-assembly to reduce reference bias. I re-wrote this software to modularize it and speed it up.

ConGen workshops

<https://gwct.github.io/congen/>

- I built these websites as a workshop resource for students during the Conservation Genomics Course.

Referee: Reference genome quality scores

<https://gwct.github.io/referee>

- This software uses genotype likelihoods from reads mapped back to their assembly to calculate a quality score for every position in the assembled genome.

***Drosophila* 25 species phylogeny**

<http://dx.doi.org/10.6084/m9.figshare.5450602>

- As part of a larger project, I inferred the phylogeny of 25 *Drosophila* species and published it standalone on FigShare as a resource for others to use.

GRAMPA: Gene-tree Reconciliation Algorithm with MUL-trees for Polyploid Analysis

<https://gwct.github.io/grampa.html>

- Given a singly-labeled species topology and a set of corresponding gene-trees, this software can infer if any whole genome duplications have occurred and, if so, infer the mode of polyploidization and the placement on the phylogeny.

i5K Phylogenomics Website

<https://arthrofam.org>

- With the vast amount of data involved in the i5K pilot project, I developed this website to organize and share the phylogenetic and comparative results with colleagues.

caferror

<https://hahnlab.github.io/CAFE/>

- Part of CAFE version 3, I wrote this program to use CAFE's error modeling function to estimate genome assembly and annotation error. The algorithm is now integrated in CAFE version 5.